

ABSTRACT OF THE DISCLOSURE

Legibility of text in an image of a page is determined by comparing a measure of the text in the page image with a measure of the page image itself. In one aspect, a measure of the text in the page image may be the height of a line of text, while the
5 measure derived from the page image may be the height of the page image. A text-to-page height ratio is determined and compared to one or more thresholds for determining legibility. In another aspect of the invention, a measure of the text in a page image is obtained by measuring the word density in the page image, while the measure derived from the page image comprises compressing the page image and determining the
10 size of the compressed image file. Legibility is then determined by comparing the measure of word density with the compressed image file size.